DOCUMENT RESUME

ED 266 686 FL 015 542

AUTHOR de Jong, John H. A. L.
TITLE Effect of Mixing Ability Groups on Ability Levels Attained.
PUB DATE 85
NOTE 11p.; In: Language Testing in School. AFinLA Yearbook 1985. No. 41 (see FL 015 537). Document is in small print.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Ability; Academic Achievement; Comparative Analysis; *Curriculum Design; *Educational Change; English (Second Language); Foreign Countries; German; *Heterogeneous Grouping; Language Tests; *Program Effectiveness; Secondary Education; *Second Language Learning
IDENTIFIERS *Netherlands

ABSTRACT

In 1983, in the Netherlands' highly differentiated school system, two types of curriculum representing different ability levels were combined as a first step towards a more heterogeneous grouping of student abilities. A study of one aspect of the results of this change compared over 1000 samples of English and German second language listening comprehension tests administered before and after the change. Statistical analysis of test results from students representing two different ability levels in different years indicate that the curriculum change may have had a negative effect at both levels of achievement. The results indicate that mixing ability levels has a negative effect on the final ability attained. The findings are supported by results on tests of other foreign language skills (reading comprehension) and in another subject (chemistry). (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Tomasla

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED266686

EFFECT OF MIXING ABILITY GROUPS ON ABILITY LEVELS ATTAINED

JOHN H.A.L. de Jong (Cito, Netherlands)

Summary

In the Netherlands the system of secondary education is highly differentiated. Possibly as a first step towards a comprehensive system two different types of secondary education were brought together in 1983. However, because of the width of the ability range, examinations at two different levels are offered at the end of the curriculum. The two levels were meant to be equivalent to those in the two types of secondary education of the former system. Data from samples of 300 to 1000 examinees on English and German foreign language listening comprehension tests were collected before and after the introduction of the new system. In an equating procedure based on Item Response Theory, using the Rasch Model, tests of different levels and from different years are interlinked by means of common items and common representative samples. The change in the educational system is found to have had a negative effect at both levels of achievement. These findings are corroborated by results on tests of other foreign language skills (reading comprehension) and other subjects (chemistry).

Introduction

Secondary education in the Netherlands is a highly differentiated system. Four different schooltypes aim at four different proficiency levels.

LBO provides lower vocational training and is attended by about 22% of the age cohort. LBO has a four year programme. Final examinations are taken as a rule at the age of 16. The foreign language programme takes about 300 hours and leads to a level comparable to the Threshold Level as defined by the Council of Europe (Van Ek, 1975).

MAVO is the lowest level of general education, meant as a basis for further training for clerical jobs and is attended by about 37% of the age cohort. MAVO has a four year programme. Final examinations are taken at the age of 16. The foreign language programme takes about 400 hours.

HAVO is an intermediate level of general education, meant to prepare for further studies leading to higher non-academic jobs and is attended by about 22% of the age cohort. Final examinations are at the age of 17. The foreign language programme takes about 500 hours and the final level is comparable to that reached by secondary education in most western countries.

BEST COPY AVAILABLE

FL015542

VWO is the highest level of general education meant as a preparation for academic studies and is taken by about 18% of the age cohort. VWO is a six year programme. Final examinations are taken at the age of 18. The foreign language programme takes about 600 hours.

Within the MAVO schools, however, two subtypes have to be distinguished. Until 1979 pupils could choose either a three or a four year programme: MAVO-3 and MAVO-4. Final examinations for MAVO-3 were at a level below the regular (i.e.: MAVO-4) level. This MAVO-3 level was comparable to the LBO level and was taken by about 5% of all pupils attending MAVO. For several school subjects examination papers were identical for LBO and MAVO-3. In 1979 it was decided to bring all MAVO pupils together in a single school with the same four year curriculum for all. At the end of the curriculum pupils have the choice - for each subject separately - to take examinations at two different levels: MAVO-C or MAVO-D. These levels are meant to correspond to the MAVO-3 and MAVO-4 levels in the former system and several examination papers are once again identical for LBO and MAVO-C. The rationale behind the change of the system for MAVO is to offer all students better chances. A MAVO-3 certificate was not greatly appreciated. It was therefore felt to be better for MAVO students to be able to postpone the definitive choice for a particular level; to offer all MAVO students the opportunity to obtain the higher level MAVO certificate. If this aim should prove to be too high for some, then the possibility of taking exams in a number of subjects at the lower level would give them a chance to obtain a certificate with the prestige of the higher level MAVO for at least the remainder of the subjects.

In 1983 the first cohort of MAVO pupils in the new system presented themselves for final examinations. The present study was undertaken to reveal the effect of the changes in the MAVO schools on the achievement level of MAVO pupils: did the postponement of level determination and the lengthening of the curriculum for the lower level lead to better results and to higher certification?

Method

The method used is based on Item Response Theory using the Rasch Model (Rasch, 1960). In the Rasch Model one item parameter and one person parameter determine the probability that a given person answers a given item correctly. The item parameter in the Rasch Model is called the difficulty parameter and the person parameter is called the ability parameter. However, since for the calibration of each test an arbitrary origin of measurement is selected the numerical values of the parameters of different tests cannot be compared directly. In order to make them comparable a common origin has to be determined for the different tests. This procedure, known as test equating, can be followed if:

- the tests contain a number of common items;
- the groups taking the tests contain common persons;

BEST COPY AVAILABLE

- the groups taking the tests are representative samples from the same population.

If one of these conditions is met, and if the tests measure the same ability, and the tests are not to far apart in difficulty and reliability (Petersen et al., 1982) the item parameters of the tests and the ability parameters of the persons can be equated, i.e: the parameters can be put on a single scale.

In 1982 (the last year of MAVO examinations in the former system) random samples of 300 were taken from the LBO, MAVO-3 and MAVO-4 populations. Because the LBO and MAVO-3 levels are supposed to be identical LBO and MAVO-3 students take the same foreign language listening comprehension test. The students in each sample were asked to take one third of the test of the other level besides taking the test at their own level: thus LBO and MAVO-3 students took a part of the MAVO-4 test and MAVO-4 students took a part of the LBO/MAVO-3 test. In 1983 and 1984 similar procedures were followed with samples of 300 to 1000 from the LBO, MAVO-C and MAVO-D populations. This design would permit equating both the tests and the populations concerned by means of a common item link. Random samples of 1000 LBO pupils from the years 1982, 1983 and 1984 were taken to form representative samples of the same population, thus permitting to equate the scales of the subsequent years.

The tests used in this study are the regular Cito tests of foreign language listening comprehension. The tests were originally developed in a research project at the University of Utrecht (Groot, 1975). Because examinations are public, new tests have to be developed each year. Cito tests of foreign language listening comprehension have been constructed at Cito and used in final examinations in schools for over ten years. The objective of the tests, as defined by Groot (1975), is to evaluate foreign language learners' ability to understand the foreign language, spoken spontaneously by educated native speakers at normal conversational speed. Extremely informal elements as well as lexical and syntactical elements that are incomprehensible to less educated native speakers and topics requiring specific knowledge are excluded from the tests. The language material consists of three or four interviews with native speakers of various occupations and professions. The item format is a multiple choice question with three options. The language material is divided in samples of forty to fifty seconds; the correct answer is a one phrase summary of the global contents of the sample. Each sample is followed by a pause of twenty seconds on the tape which allows the students to read the item in their test book and tick their answer on an answer sheet. A second item format was introduced (De Jong, 1983; De Jong, 1984): modified cloze items with two options. In each sample of twenty to thirty seconds on the tape a word, or group of words is deleted and replaced by an electronic signal. Words to be deleted from the original text are chosen for their semantic relevance in the context. The language material consists of discussions, conversations, news items and other radio programmes. The students are to decide in a pause of seven seconds

following the language samples on tape, which of the two options in their test book can be used to restore the original text. Depending on level and item format total test length may vary from forty to fifty items. Administration time varies between fifty and sixty minutes. Test reliability (KR20) ranges from .75 to .85.

The test analyses according to the Rasch Model were performed with the computer programme CALFIT (Wright and Mead, 1975). Tests were equated following to the procedure described by Wright and Stone (1979).

Results and discussion

Figure 1 presents the results from the equating procedures on the German foreign language listening comprehension tests for HAVO in the years 1982, 1983 and 1984. The three horizontal lines in figure 1a represent the ability variables in the subsequent years, but should in fact be regarded as a single line as the tests have been calibrated on a single scale. Above the lines the mean ability at the higher HAVO level is indicated by triangles. The shaded areas represent the distribution of ability from two standard deviations below the mean to two standard deviations above the mean, thus covering about 96% of the population. Arrows indicate the cut-off point. Below each line the same is shown for the lower HAVO level. The dot on each line represents the LBO mean ability in the subsequent years, the dots are connected by a dotted vertical line. The figure shows that there is a difference of about one logit (the unity of measurement on the ability scale) between the two HAVO levels. This difference remains fairly constant during the three years. The mean ability of both HAVO levels shifts to the left, that is: to the lower end of the ability scale in 1983. Although the mean ability at both levels slightly moves to the right again in 1984 it does not reach the original level of 1982.

The number of students at the two HAVO levels is not equal and varies from year to year. A larger proportion of the total HAVO population has taken examinations at the lower HAVO level from 1983 on. This has an effect on the mean ability of the total HAVO population. The mean ability and the distribution of ability in the total HAVO population is pictured in Figure 1b. Figure 1b is drawn at the same scale as Figure 1a. The arrows indicate the mean ability in the subsequent years; the shaded area represents the distribution of ability of 96% of the total HAVO population (both levels). Apart from the shift to the lower end of the ability scale a smaller spread in ability can be observed.

Figures 2a and 2b picture the same kind of data for English listening comprehension. Figure 2a shows a comparable shift to the lower end of the ability scale as was found for German. For English, however, the mean ability of the lower HAVO level, which was well above the mean LBO level in 1982, is below LBO in 1983. Again, as for German, a slight shift to the right in 1984 does not result in

complete recovery of the original level. The combination of the two MAVO levels in Figure 2b shows a similar narrowing of the spread in ability of the total MAVO population as was found for German.

An observed shift in mean ability of a population of about 90,000 students cannot be interpreted as a standard error of the mean. Furthermore, it is highly improbable that such a standard error of the mean would cause deviation in the same direction for two different years and for two different languages. To check on sampling errors reference was made to the results of the populations concerned on reading comprehension tests. Results on the reading comprehension tests are collected each year from the total population: 40,000 LBO students and 90,000 MAVO students. Figure 3 pictures the relative position of the lower level MAVO students and the LBO students from 1978 to 1984 in standard deviations of test scores. The mean LBO score is set at zero, the horizontal axis represents time and the vertical axis standard deviations of test scores on the reading comprehension tests. The mean score of the lower level MAVO students for English reading comprehension was traditionally well above the mean score of LBO students. This situation changes in 1983: the lower level MAVO students drop to a level that is more than half a standard deviation below their former level. For German a similar drop can be observed, in this case enlarging the original distance between LBO and lower level MAVO.

Another possible influence on the mean ability of the MAVO and LBO students could be a change in the groups involved. If quite a large proportion of the age cohort has chosen for a different schooltype since the change of the system, a shift in mean ability of the groups is to be expected. Figure 4 pictures the number of examinees taking English (ca. 97% of the total population) in the different schooltypes from 1972 to 1984. Looking at the period from 1972 to 1982 a constant growth, parallel to the postwar increase in population can be observed. The two MAVO groups are an exception in that the lower level MAVO group decreases from 1980 on going with a more rapid growth of the higher level MAVO. Immediately after the change of the MAVO system a dramatic change in the two MAVO populations can be observed: more than three times as many MAVO students opt for the lower level MAVO. There is no indication of a shift in population between LBO and MAVO. The larger number of lower level MAVO students is apparently recruited from the group that formerly opted for the higher level MAVO and would therefore be expected to raise the level of achievement in the lower level group. Furthermore, hypothesizing that it will not have been the highest achievers in the higher level group that deserted this level for the lower, one could equally have expected a rise in mean ability for the higher level MAVO. The observed change in numbers of the MAVO groups therefore, cannot account for the downward shift in achievement. On the contrary, this change would lead one to expect a shift in the opposite direction. Bakker (1984) reports that the hypothesis of a rise in mean ability of the higher MAVO group in chemistry, due to the changed distribution of MAVO students over the two MAVO levels

BEST COPY AVAILABLE

after the change of the system, has to be rejected.

Apart from the effect on the achievement level of MAVO students the change in the MAVO system has had its effect on their level of certification. Figure 5 pictures this effect. The proportion of MAVO students taking a subject at the higher MAVO level has gone down for all subjects. Figure 5a shows the exact proportions in 1982 and 1983 for English. For the foreign languages the proportion of ca. .05 of the total MAVO population that took the lower MAVO level has gone up to a proportion of .15 to .18. For subjects such as mathematics and chemistry, traditionally regarded as difficult, the original .05 may go up to proportions of .20 to .40.

The advocates of the change in the MAVO system present the statistics of Figure 5b as proof of a positive effect. Before the change of the system 15% to 20% of the total MAVO population finished their last year without any certificate at all and had to sit the last year again to obtain a certificate, which meant the loss of an entire year for these students. In the new system, only about 7% or 8% fail to obtain a certificate. However, one whole year has been added to the curriculum for about 5% of the students: they have had their "loss of a year".

To make a fair comparison between the certification of MAVO students before and after the change of the system, the level of certification also has to be taken into account. Figure 5c shows that the proportion of Mavo students taking all subjects at the lower level has remained more or less the same in 1982 and 1983. The number of failures amongst this group has increased dramatically: from about 20% in 1982 to almost 50% in 1983. The proportion of candidates taking MAVO exams in all subjects at the higher level in 1983 is only half of that of the years before. Granted, the students taking all subjects at the higher level do so with success in 1983 a larger proportion passes the exams, but it is much smaller group than before. A smaller proportion of MAVO students will in effect enjoy the greater social appreciation associated with the higher level certificate. The proportion of MAVO students passing at the higher level MAVO for all subjects has gone down from about .80 in 1982 to about .53. Between lower level MAVO and higher level MAVO intermediate levels have been created; students can now have mixed level certificates. These mixed certificates can contain one subject at the lower level and five subjects at the higher level; or five subjects at the lower level and one subject at the higher level or any other combination with a total of six subjects. How exactly the mixed level certificates will be appreciated cannot be foreseen, but a certificate with all subjects taken at the higher level will no doubt receive more appreciation.

After only two years it is too early to pronounce a final judgement on the new MAVO system. However, the downward shift in the mean ability of the total MAVO population for listening comprehension of German and English, the relatively lower scores for the lower level MAVO on reading comprehension tests of German and English, the absence of a rise in mean ability in chemistry in spite of a more

select group, the smaller proportion of MAVO students obtaining a certificate with all subjects taken at the higher level, all these observations reveal a similar tendency: mixing ability groups has a negative effect on the final ability attained. Various explanations for this negative effect can be offered. Teachers may tend to adapt their level of teaching to the mean level of the students in the group. Students may be less motivated to do their utmost when they know failure at the higher level does not mean total failure: they can always try their success at a lower level. Though more data on the results of the new system are needed, it is impossible to welcome the present results and extreme caution has to be taken in further experiments towards a comprehensive system.

References

Bakker, S. 1984. MAVO-niveau op drift. In: Weekblad voor leraren van het M.G.L. 17, 8, 298-299.

De Jong, J.H.A.L. 1983. Focusing in on a latent trait; An attempt at construct validation by means of the Rasch model. In: J. van Weeren (ed.), Practice and problems in language testing 5: Non-classical test theory. Final examinations in secondary schools. National Institute for Educational Measurement, Cito, Arnhem.

De Jong, J.H.A.L. 1984. Testing foreign language listening comprehension. In: Language Testing 1,1,97-100.

Groot, P.J.M. 1975. Testing communicative competence in listening comprehension. In: R.L. Jones and B. Spolsky (eds.), Testing language proficiency. Centre for Applied Linguistics, Arlington, Virginia.

Petersen, N.S. Marco, G.L. and Stewart, E.E. 1982. A test of the adequacy of linear score equating models. In: P.W. Holland and D.B. Rubin (eds.), Test equating. Academic Press, New York.

Rasch, G. 1960. Probabilistic models for some intelligence and attainment tests. Danmarks Pædagogiske Institut, København.

Van Ek, J.A. 1975. The threshold level. With an appendix by L.J. Alexander. Council of Europe, Strasbourg.

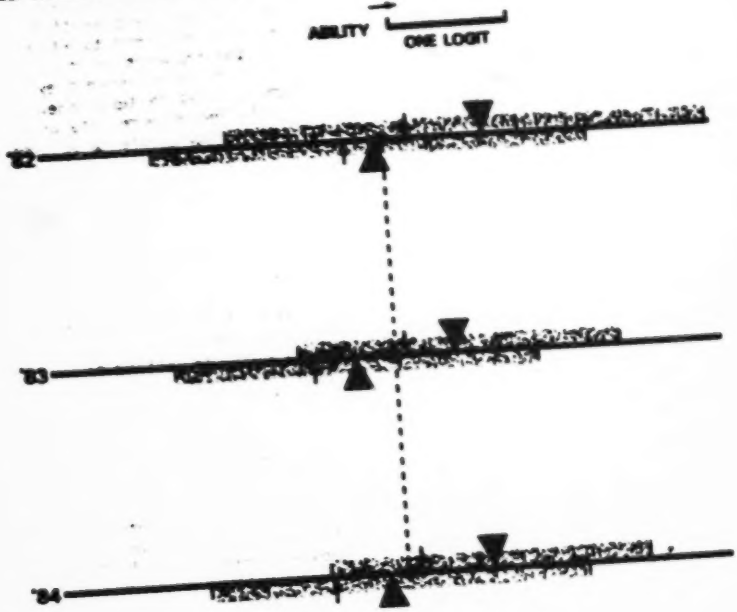
Wright, B.D. and Mead, R.J. 1975. CALFIT: Research Memorandum nr. 18. Department of Education, University of Chicago, Chicago.

Wright, B.D. and Stone M.H. 1979. Best test design: Rasch measurement. Mesa Press, Chicago.

BEST COPY AVAILABLE

Figure 1 German listening comprehension 1982-1984 :
Mean ability and distribution of ability of MAVO candidates

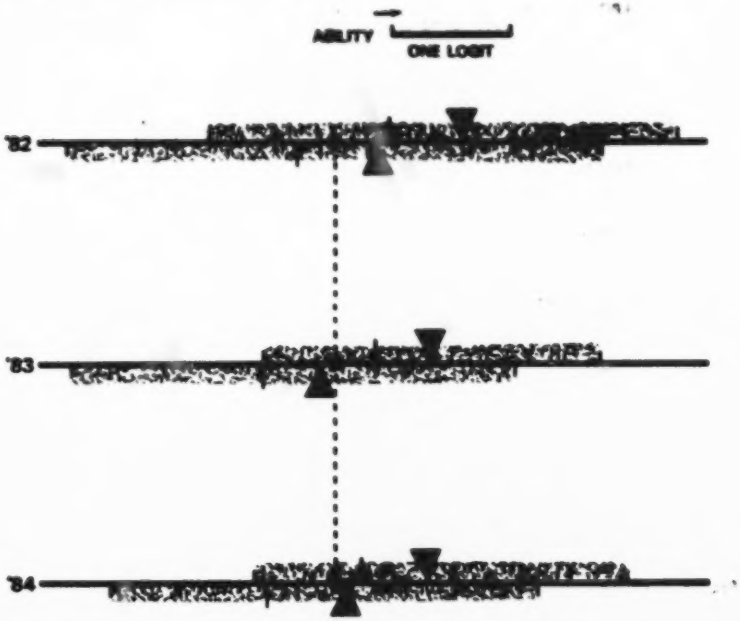
1a At the MAVO levels separately



1b Both MAVO levels combined

Figure 2 English listening comprehension 1982-1984 :
Mean ability and distribution of ability of MAVO candidates

2a At two MAVO levels separately



2b Both MAVO levels combined

Figure 3 Reading comprehension 1978-1984 :
Relative achievement level of LBO and MAVO candidates

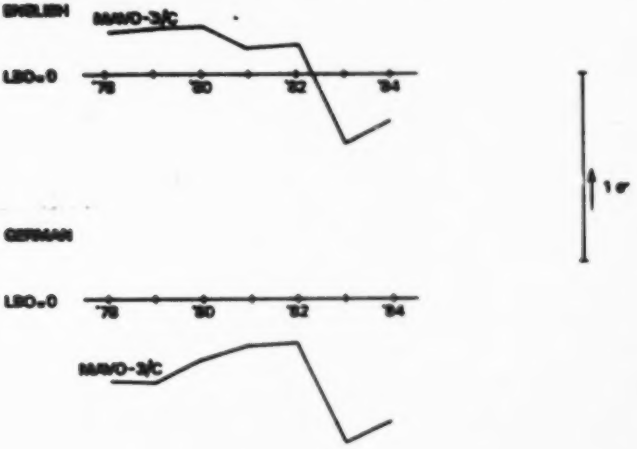


Figure 4 Number of examinees taking English;per schooltype 1976-1984

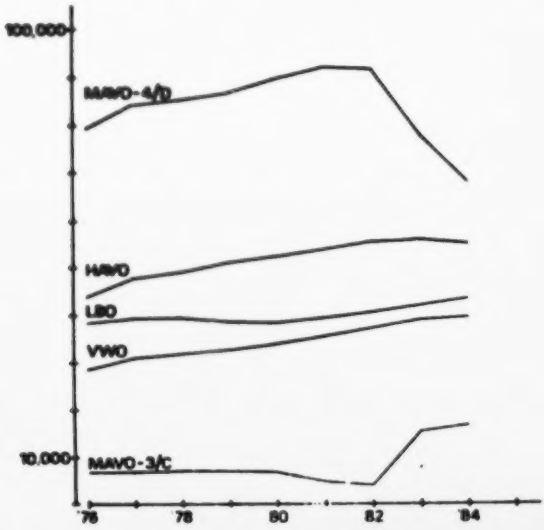
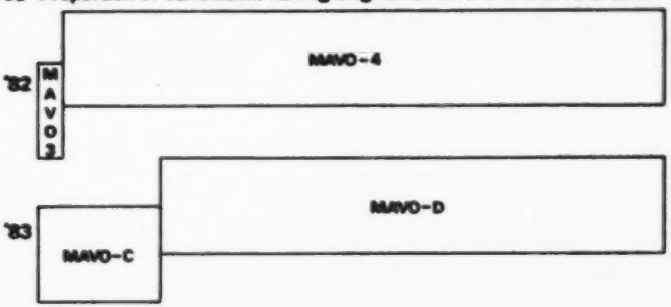
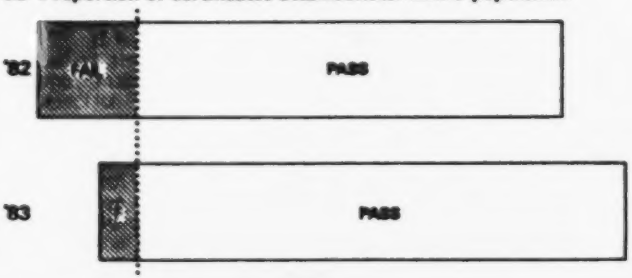


Figure 5 MAVO examination levels attained in 1982 and 1983

5a Proportion of candidates taking English at MAVO 3/C or 4/D level



5b Proportion of certificates attained:total MAVO population



5c Proportion of certificates attained:per MAVO level